A LOCALIZED ENSEMBLE KALMAN SMOOTHER

*Mark D. Butala**

Jet Propulsion Laboratory, California Institute of Technology
National Aeronautics and Space Administration

ABSTRACT

Numerous geophysical inverse problems prove difficult because the available measurements are indirectly related to the underlying unknown dynamic state and the physics governing the system may involve imperfect models or unobserved parameters. Data assimilation addresses these difficulties by combining the measurements and physical knowledge. The main challenge in such problems usually involves their high dimensionality and the standard statistical methods prove computationally intractable. This paper develops and addresses the theoretical convergence of a new high-dimensional Monte Carlo approach called the localized ensemble Kalman smoother.

***Index Terms*—** multidimensional signal processing; recursive estimation; Kalman filter; remote sensing

1. INTRODUCTION

The ensemble Kalman filter (EnKF) [1] has seen extensive application to a wide variety of high-dimensional data assimilation applications (see references in [1]), including oceanography, operational use in numerical weather prediction, and solar imaging [2]. The goal in data assimilation is to combine empirical measurements with a first-principles dynamic model to produce a statistically optimal time-dependent solution to the underlying high-dimensional inverse problem. The popularity of the EnKF is due to the relative simplicity of its implementation, ease of parallelization, no requirement for adjoint physics models which can prove difficult to derive, and competitive performance relative to other data assimilation approaches.

Often in data assimilation, the filtered or causal estimate is of interest. Forecasts are usually derived by applying the dynamic model to the most recent filtered estimate to predict the future physical state. When the dynamics are chaotic or incompletely modeled and forecasting is difficult or even impossible, then the filtered estimate can serve as a useful "nowcast" of the current physical state. Smoothed estimates utilize all available measurement to produce retrospective noncausal estimates of the physical state with the least error.

*The publication of this work was supported by JPL under a contract to NASA.

Several ensemble Kalman smoother (EnKS) approaches have been developed to solve smoothing problems in data assimilation. One approach is given in [1], but the method is limited to relatively low-dimensional problems. The approach most similar to this work is by Khare et al. [3], which considers high-dimensional applications. However, they do not consider the convergence properties of their approach and neglect the smoothed error covariance.

In this paper, we present a new EnKS that utilizes localization to reduce the computational cost for high-dimensional applications. Localization has been used extensively for this purpose for the EnKF and works by imposing an a priori physically motivated correlation length restriction on the structure of the error covariance matrix. The resultant Monte Carlo approximated error covariance is then guaranteed to be sparse and more easily handled in subsequent computations. The theoretical contributions of this work include a theorem regarding the convergence of the method as the ensemble size increases and a connection to a result known in radar processing that provides the means to implement a numerically stable square-root form [4] localized EnKF or EnKS.

The remainder of the paper is organized as follows. Section 2 defines the linear state-space signal model considered in this work. Section 3 provides the equations for the Kalman filter (KF) that serves as the first step for the Bryson-Frazier (BF) Kalman smoother (KS) defined in Section 4. Next, the localized EnKS is developed in Section 5 and is shown to be a localized Monte Carlo implementation of the BF KS by the convergence theorem in Section 6. A numerical example is provided in Section 7 showing a comparison between the KS, localized KS, and EnKS. Finally, conclusions are given in Section 8.

2. SIGNAL MODEL

The linear dynamic signal model is defined by the state-space equations:

$$\boldsymbol{x}_{i+1} = \boldsymbol{F}_i \boldsymbol{x}_i + \boldsymbol{u}_i, \qquad \boldsymbol{y}_i = \boldsymbol{H}_i \boldsymbol{x}_i + \boldsymbol{v}_i. \tag{1}$$

In the above, the N-dimensional vector $\boldsymbol{x}_i$ is the unknown at time index i, the time index has the range $1 \leq i \leq I$, the M-dimensional vector $\boldsymbol{y}_i$ is the ith measurement, the matrices $\boldsymbol{F}_i$ and $\boldsymbol{H}_i$ are known, as are the covariances $\boldsymbol{Q}_i \triangleq \mathrm{Cov}(\boldsymbol{u}_i)$,

$\boldsymbol{R}_i \triangleq \text{Cov}(\boldsymbol{v}_i)$, and $\boldsymbol{\Pi}_1 \triangleq \text{Cov}(\boldsymbol{x}_1)$ and initial state mean $\boldsymbol{\mu_1} \triangleq \mathbb{E}(\boldsymbol{x}_1)$.

3. KALMAN FILTER

Under the signal model of Section 2, the well-known KF [4] is the method to compute $\widehat{\boldsymbol{x}}_{i|i}$, the filtered linear minimum mean square error (LMMSE) estimate of the unknown state $\boldsymbol{x}_i$ given the set of measurements $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_i\}$. The method proceeds recursively in two stages. The first is the measurement update defined by

$$\boldsymbol{K}_i = \boldsymbol{P}_{i|i-1}\,\boldsymbol{H}_i^T\,(\boldsymbol{H}_i\,\boldsymbol{P}_{i|i-1}\,\boldsymbol{H}_i^T + \boldsymbol{R}_i)^{-1} \tag{2}$$

$$\widehat{\boldsymbol{x}}_{i|i} = \widehat{\boldsymbol{x}}_{i|i-1} + \boldsymbol{K}_i\,(\boldsymbol{y}_i - \boldsymbol{H}_i\,\widehat{\boldsymbol{x}}_{i|i-1}) \tag{3}$$

$$\boldsymbol{P}_{i|i} = \boldsymbol{P}_{i|i-1} - \boldsymbol{K}_i\,\boldsymbol{H}_i\,\boldsymbol{P}_{i|i-1} \tag{4}$$

where $\boldsymbol{K}_i$ is commonly referred to as the Kalman gain. The second step is the time update:

$$\widehat{\boldsymbol{x}}_{i+1|i} = \boldsymbol{F}_i\,\widehat{\boldsymbol{x}}_{i|i}, \qquad \boldsymbol{P}_{i+1|i} = \boldsymbol{F}_i\,\boldsymbol{P}_{i|i}\,\boldsymbol{F}_i^T + \boldsymbol{Q}_i. \tag{5}$$

4. BRYSON-FRAZIER KALMAN SMOOTHER

The BF KS [4] consists of three stages. The first stage processes the complete data set through the KF described in Section 3 and, for each time index i, stores the following quantities to disk for later use: the Kalman gain $\boldsymbol{K}_i$ (2), the innovation $\boldsymbol{e}_i \triangleq \boldsymbol{y}_i - \boldsymbol{H}_i\,\widehat{\boldsymbol{x}}_{i|i-1}$, and the innovation covariance

$$\boldsymbol{R}_{e,i} \triangleq \text{Cov}(\boldsymbol{e}_i) = \boldsymbol{R}_i + \boldsymbol{H}_i\,\boldsymbol{P}_{i|i-1}\,\boldsymbol{H}_i^T \tag{6}$$

in addition to the filtered estimate $\widehat{\boldsymbol{x}}_{i|i}$ and error covariance $\boldsymbol{P}_{i|i}$. The second stage processes the time-reversed data set to compute the adjoint variable (where $\boldsymbol{\lambda}_{I+1} = \boldsymbol{0}$)

$$\boldsymbol{\lambda}_i = (\boldsymbol{I} - \boldsymbol{K}_i\,\boldsymbol{H}_i)^T\,\boldsymbol{F}_i^T\,\boldsymbol{\lambda}_{i+1} + \boldsymbol{H}_i^T\,\boldsymbol{R}_{e,i}^{-1}\,\boldsymbol{e}_i \tag{7}$$

and its covariance (where $\boldsymbol{\Lambda}_{I+1} = \boldsymbol{0}$, the matrix with all elements equal to 0)

$$\boldsymbol{\Lambda}_i = (\boldsymbol{I} - \boldsymbol{K}_i\,\boldsymbol{H}_i)^T\,\boldsymbol{F}_i^T\,\boldsymbol{\Lambda}_{i+1}\,\boldsymbol{F}_i\,(\boldsymbol{I} - \boldsymbol{K}_i\,\boldsymbol{H}_i) + \boldsymbol{H}_i^T\,\boldsymbol{R}_{e,i}^{-1}\,\boldsymbol{H}_i \tag{8}$$

through the above backwards recursions. The final stage combines quantities found in the first two stages to compute the smoothed estimate and error covariance:

$$\widehat{\boldsymbol{x}}_{i|1:I} = \widehat{\boldsymbol{x}}_{i|i} + \boldsymbol{P}_{i|i}\,\boldsymbol{F}_i^T\,\boldsymbol{\lambda}_{i+1} \tag{9}$$

$$\boldsymbol{P}_{i|1:I} = \boldsymbol{P}_{i|i} - \boldsymbol{P}_{i|i}\,\boldsymbol{F}_i^T\,\boldsymbol{\Lambda}_{i+1}\,\boldsymbol{F}_i\,\boldsymbol{P}_{i|i}. \tag{10}$$

5. ENSEMBLE KALMAN SMOOTHER

The EnKS can be thought of as a Monte Carlo approximation to the BF KS that utilizes localization. The first stage of the EnKS is the EnKF (see [5] for a development that uses the same notation as this work). At each time index i, the following quantities are stored to disk for later use: the filtered estimate $\widetilde{\boldsymbol{x}}_{i|i}$, the ensemble Kalman gain $\widetilde{\boldsymbol{K}}_i$ (see (23) in [5]), the ensemble innovation $\widetilde{\boldsymbol{e}}_i \triangleq \boldsymbol{y}_i - \boldsymbol{H}_i\,\widetilde{\boldsymbol{x}}_{i|i-1}$, the ensemble innovation covariance

$$\widetilde{\boldsymbol{R}}_{e\,i} \triangleq \boldsymbol{R}_i + \boldsymbol{H}_i\,(\boldsymbol{C}_i \circ \widetilde{\boldsymbol{P}}_{i|i-1})\,\boldsymbol{H}_i^T, \tag{11}$$

where $\boldsymbol{C}_i$ is the covariance taper matrix that defines the a priori error covariance localization structure, $\circ$ denotes the Hadamard or element-by-element matrix product. The prior and posterior ensembles denoted $\widetilde{\boldsymbol{X}}_{i|i-1}$ and $\widetilde{\boldsymbol{X}}_{i|i}$, respectively, must also be stored where the lth column of the matrix $\widetilde{\boldsymbol{X}}_{i|j}$ is given by $\left[\widetilde{\boldsymbol{X}}_{i|j}\right]_{(:,\,l)} = \widetilde{\boldsymbol{x}}_{i|j}^{\,l}$ where $\widetilde{\boldsymbol{x}}_{i|j}^{\,l}$ is the lth ensemble member.

The second stage of the EnKS involves the following backwards recursion on the ensemble adjoint variable (where $\widetilde{\boldsymbol{\lambda}}_{I+1} = \boldsymbol{0}$):

$$\widetilde{\boldsymbol{\lambda}}_i = \left(\boldsymbol{I} - \widetilde{\boldsymbol{K}}_i\,\boldsymbol{H}_i\right)^T\boldsymbol{F}_i^T\,\widetilde{\boldsymbol{\lambda}}_{i+1} + \boldsymbol{H}_i^T\,\widetilde{\boldsymbol{R}}_{e,\,i}^{-1}\,\widetilde{\boldsymbol{e}}_i \tag{12}$$

and its covariance (where $\widetilde{\boldsymbol{\Lambda}}_{I+1} = \boldsymbol{0}$)

$$\widetilde{\boldsymbol{\Lambda}}_i = \left(\boldsymbol{I} - \widetilde{\boldsymbol{K}}_i\,\boldsymbol{H}_i\right)^T\boldsymbol{F}_i^T\,\widetilde{\boldsymbol{\Lambda}}_{i+1} + \widetilde{\boldsymbol{R}}_{e,\,i}^{-1/2}\,\boldsymbol{Z}_i \tag{13}$$

and $[\boldsymbol{Z}_i]_{(:,\,l)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{0},\,\boldsymbol{I})$. It is easily shown that $\boldsymbol{\Lambda}_i \approx \widetilde{\boldsymbol{\Lambda}}_i\,\widetilde{\boldsymbol{\Lambda}}_i^T/(L-1)$ and the approximation is within the sample error when no covariance tapering is applied, i.e., when $\boldsymbol{C}_i = \boldsymbol{1}$ (a matrix with all elements equal to 1).

The presence of a square root of the ensemble innovation covariance $\widetilde{\boldsymbol{R}}_{e,\,i}^{1/2}$ in (13) poses a significant implementation challenge especially. Any hope to find a square root factor to (11) rests on factoring the prior tapered ensemble covariance. To find this square root factor, first note that

$$\boldsymbol{C}_i \circ \widetilde{\boldsymbol{P}}_{i|i-1} = \left(\boldsymbol{C}_i^{1/2}\,\boldsymbol{C}_i^{T/2}\right) \circ \left(\widetilde{\boldsymbol{P}}_{i|i-1}^{1/2}\,\widetilde{\boldsymbol{P}}_{i|i-1}^{T/2}\right) \tag{14}$$

where $\boldsymbol{A}^{T/2} \triangleq \left(\boldsymbol{A}^{1/2}\right)^T$, $\boldsymbol{C}_i^{T/2}$ is a convolution matrix used in the construction of the covariance taper matrix, and $\widetilde{\boldsymbol{P}}_{i|i-1}^{1/2} = \widetilde{\boldsymbol{X}}_{i|i-1}/(\sqrt{L-1})$. Square root factors involving Hadamard products such as (14) have been addressed in the literature on spatial temporal adaptive processing in radar applications [6]. The following lemma addresses the relevant square root factorization.

Lemma 1: Let $\boldsymbol{A}$ be an $M \times N$ matrix and $\boldsymbol{B}$ be an $M \times P$ matrix. Then

$$\left(\boldsymbol{A}\,\boldsymbol{A}^T\right) \circ \left(\boldsymbol{B}\,\boldsymbol{B}^T\right) = \left(\boldsymbol{A}^T \odot \boldsymbol{B}^T\right)^T\left(\boldsymbol{A}^T \odot \boldsymbol{B}^T\right) \tag{15}$$

where $\odot$ denotes the Khatri-Rao matrix product. The Khatri-Rao product of the matrices $\boldsymbol{A}^T$ and $\boldsymbol{B}^T$ is defined by

$$\boldsymbol{A}^T \odot \boldsymbol{B}^T = \begin{bmatrix} [\boldsymbol{A}]_{(1,\,:)} \otimes [\boldsymbol{B}]_{(1,\,:)} & \cdots & [\boldsymbol{A}]_{(M,\,:)} \otimes [\boldsymbol{B}]_{(M,\,:)} \end{bmatrix} \tag{16}$$

where $\otimes$ is the Kronecker matrix product, $[\cdot]_{(m,:)}$ selects the mth row of the matrix argument, and $A^T \odot B^T$ has dimensions $NP \times M$. (See the proof to P 6.4.2 in [7])
The square root factorization of (11) also requires the following lemma.

Lemma 2: Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $M \times N$ ($M \leq N$) matrices. Then $\boldsymbol{A}\,\boldsymbol{A}^T = \boldsymbol{B}\,\boldsymbol{B}^T$ if, and only if, there exists an $N \times N$ unitary matrix $\boldsymbol{\Theta}$ such that $\boldsymbol{A} = \boldsymbol{B}\,\boldsymbol{\Theta}$. (See the proof to Lemma A.5.1 in [4])
Finally, the square root factorization of the ensemble innovations covariance (11) required in (13) is summarized in the following theorem.

Theorem 3: Let $\boldsymbol{\mathcal{Q}}_i$ and $\boldsymbol{\mathcal{R}}_i$ be the unitary and the upper triangular matrices found in the QR decomposition of the matrix

$$\boldsymbol{A}_i = \left[\boldsymbol{H}_i\big(\boldsymbol{C}_i^{T/2} \odot \widetilde{\boldsymbol{P}}_{i|i-1}^{T/2}\big)^T \quad \boldsymbol{R}_i^{1/2}\right]^T. \tag{17}$$

Then, $\boldsymbol{R}_{e,\,i}^{1/2} = [\boldsymbol{\mathcal{R}}_i]_{(1:M,\,:)}^T$ where $[\cdot]_{(1:M,\,:)}$ selects the first M rows of its matrix argument.

Proof. First note that $\boldsymbol{A}_i^T\,\boldsymbol{A}_i = \left[\widetilde{\boldsymbol{R}}_{e,i}^{1/2} \quad \boldsymbol{0}\right]\left[\widetilde{\boldsymbol{R}}_{e,i}^{1/2} \quad \boldsymbol{0}\right]^T$ by Lemma 1. Then, by Lemma 2, the above implies that there exists a unitary matrix $\boldsymbol{\Theta}_i$ such that $\boldsymbol{A}_i^T = \left[\widetilde{\boldsymbol{R}}_{e,i}^{1/2} \quad \boldsymbol{0}\right]\boldsymbol{\Theta}_i$. Note that the QR decomposition provides the following factorization: $\boldsymbol{A}_i = \boldsymbol{\mathcal{Q}_i}\,\boldsymbol{\mathcal{R}_i}$. Finally, note that $\boldsymbol{A}_i^T = \boldsymbol{\mathcal{R}_i}^T\,\boldsymbol{\mathcal{Q}_i}^T$ and $\boldsymbol{\mathcal{Q}_i}^T$ is a unitary matrix. □

The third and final stage of the EnKS computes the ensemble approximation to the smoothed estimate

$$\widetilde{\boldsymbol{x}}_{i|1:I} = \widetilde{\boldsymbol{x}}_{i|i} + \big(\boldsymbol{C}_i' \circ \widetilde{\boldsymbol{P}}_{i|i}\big)\boldsymbol{F}_i^T\,\widetilde{\boldsymbol{\lambda}}_{i+1} \tag{18}$$

and the smoothed error covariance

$$\widetilde{\boldsymbol{P}}_{i|1:I} = \widetilde{\boldsymbol{P}}_{i|i} - \big(\boldsymbol{C}_i' \circ \widetilde{\boldsymbol{P}}_{i|i}\big)\,\boldsymbol{F}_i^T\,\big(\boldsymbol{C}_i'' \circ \widetilde{\boldsymbol{\Lambda}}_{i+1}\big)\,\boldsymbol{F}_i\,\big(\boldsymbol{C}_i' \circ \widetilde{\boldsymbol{P}}_{i|i}\big) \tag{19}$$

where $\boldsymbol{C}_i'$ and $\boldsymbol{C}_i''$ are two additional application dependent covariance taper matrices.

6. ENSEMBLE KALMAN SMOOTHER CONVERGENCE

The following theorem addresses the convergence of the EnKS as the ensemble size L increases. This theorem concerning the asymptotic convergence of the EnKS is important because it demonstrates that the approach converges to a well defined limit which we call the localized Kalman smoother (LKS), shows that the EnKS without covariance tapering is a Monte Carlo Bryson-Frazier smoother, and provides a means for investigating the implications of the covariance taper.

Theorem 4: For each time index i, the EnKS estimates $\widetilde{\boldsymbol{x}}_{i|1:I}$ converge in probability to the localized Kalman smoother (LKS) estimates $\widetilde{\boldsymbol{x}}_{i|1:I}^{\infty}$ defined below, i.e., $\widetilde{\boldsymbol{x}}_{i|1:I} \xrightarrow{\text{p.}} \widetilde{\boldsymbol{x}}_{i|1:I}^{\infty}$ in the limit as the ensemble size $L \to \infty$.

Proof. The proof proceeds in a manner similar to the EnKF convergence proof in Appendix III in [5]. Again, Slutsky's theorem plays an important role since the ensemble members in the EnKS are identically distributed but dependent. The proof proceeds by induction, showing that the terms in each stage converge in probability to the corresponding terms in the LKS. The details are omitted because of the close similarity to the proof from Appendix III in [5]. □

Like the BF KS, the LKS has three stages, the first being the localized KF defined by (28)-(32) in [5]. The second stage is the backwards recursion for the adjoint variable (where $\widetilde{\boldsymbol{\lambda}}_{I+1}^{\infty} = \boldsymbol{0}$)

$$\widetilde{\boldsymbol{\lambda}}_i^{\infty} = \big(\boldsymbol{I} - \widetilde{\boldsymbol{K}}_i^{\infty}\,\boldsymbol{H}_i\big)^T\boldsymbol{F}_i^T\,\widetilde{\boldsymbol{\lambda}}_{i+1}^{\infty} + \boldsymbol{H}_i^T\big(\widetilde{\boldsymbol{R}}_{e,\,i}^{\infty}\big)^{-1}\widetilde{\boldsymbol{e}}_i^{\infty} \tag{20}$$

where $\widetilde{\boldsymbol{K}}_i^{\infty}$ is the localized Kalman gain (see (28) in [5]) and

$$\widetilde{\boldsymbol{R}}_{e,\,i}^{\infty} \triangleq \boldsymbol{R}_i + \boldsymbol{H}_i\,\widetilde{\boldsymbol{P}}_{i|i-1}^{\infty}\,\boldsymbol{H}_i^T, \quad \widetilde{\boldsymbol{e}}_i^{\infty} \triangleq \boldsymbol{y}_i - \boldsymbol{H}_i\,\widetilde{\boldsymbol{x}}_{i|i-1}^{\infty}. \tag{21}$$

The backwards recursion for the covariance of the adjoint variable is given by (where $\widetilde{\boldsymbol{\Lambda}}_{I+1}^{\infty} = \boldsymbol{0}$)

$$\begin{aligned}\widetilde{\boldsymbol{\Lambda}}_i^{\infty} &= \big(\boldsymbol{I} - \widetilde{\boldsymbol{K}}_i^{\infty}\,\boldsymbol{H}_i\big)^T\boldsymbol{F}_i^T\,\widetilde{\boldsymbol{\Lambda}}_{i+1}^{\infty}\,\boldsymbol{F}_i\,\big(\boldsymbol{I} - \widetilde{\boldsymbol{K}}_i^{\infty}\,\boldsymbol{H}_i\big)\\ &\quad + \boldsymbol{H}_i^T\big(\widetilde{\boldsymbol{R}}_{e,\,i}^{\infty}\big)^{-1}\boldsymbol{H}_i.\end{aligned} \tag{22}$$

The final stage combines information in the first two stages to compute the localized smoothed estimate and error covariance with

$$\widetilde{\boldsymbol{x}}_{i|1:I}^{\infty} = \widetilde{\boldsymbol{x}}_{i|i}^{\infty} + \big(\boldsymbol{C}_i' \circ \widetilde{\boldsymbol{P}}_{i|i}^{\infty}\big)\,\boldsymbol{F}_i^T\,\widetilde{\boldsymbol{\lambda}}_{i+1|1:I}^{\infty} \tag{23}$$

$$\widetilde{\boldsymbol{P}}_{i|1:I}^{\infty} = \widetilde{\boldsymbol{P}}_{i|i}^{\infty} - \big(\boldsymbol{C}_i' \circ \widetilde{\boldsymbol{P}}_{i|i}^{\infty}\big)\,\boldsymbol{F}_i^T\,\big(\boldsymbol{C}_i'' \circ \widetilde{\boldsymbol{\Lambda}}_{i+1}^{\infty}\big)\,\boldsymbol{F}_i\,\big(\boldsymbol{C}_i' \circ \widetilde{\boldsymbol{P}}_{i|i}^{\infty}\big). \tag{24}$$

The following corollary to Theorem 4 addresses the convergence of the EnKS when no covariance tapering is applied.

Corollary 5: For each time index i, the unlocalized ($\boldsymbol{C}_i = \boldsymbol{C}_i' = \boldsymbol{C}_i'' = \boldsymbol{1}$) EnKS estimates $\widetilde{\boldsymbol{x}}_{i|1:I}$ converge in probability to the LMMSE smoothed estimates $\widehat{\boldsymbol{x}}_{i|1:I}$, i.e., $\widetilde{\boldsymbol{x}}_{i|1:I} \xrightarrow{\text{p.}} \widehat{\boldsymbol{x}}_{i|1:I}$ in the limit as the ensemble size $L \to \infty$.

Proof. When no covariance tapering is applied, the LKS and BF KS of Section 4 are equivalent. Then, apply the result of Theorem 4. □

7. NUMERICAL EXAMPLE

The goal in the following experiment is to evaluate the performance of the EnKS in a low-dimensional 1-D example. The problem dimension is small enough that the EnKS state estimates can be compared to the KS and LKS estimates. We show that the bias introduced by covariance tapering in the

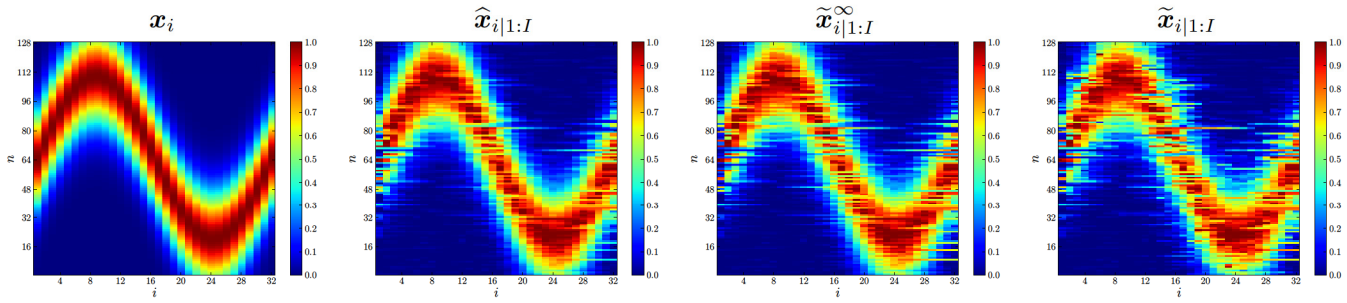

Fig. 1. The left-most image shows the ground truth state ($\boldsymbol{x}_i$) used in the harmonic oscillator numerical example described in Section 7. The horizontal axis is the time index i and the vertical axis is the nth element of $\boldsymbol{x}_i$. The remaining images show a comparison of the KS ($\widehat{\boldsymbol{x}}_{i|1:I}$), LKS ($\widetilde{\boldsymbol{x}}^{\infty}_{i|1:I}$), and EnKS ($\widetilde{\boldsymbol{x}}_{i|1:I}$) estimates.

LKS is relatively small and that the EnKS with a relatively small number of ensemble members provides estimates that are close to the LKS estimates.

The ground truth $\boldsymbol{x}_i$ is the discretized harmonic oscillator depicted in left-most image in Figure 1. The 1-D example is discretized on an $N = 128$ grid and the oscillator passes through one complete period over $I = 32$ time steps. A measurement $\boldsymbol{y}_i$ consists of $M = 64$ direct noisy observations, i.e., the matrix $\boldsymbol{H}_i$ has M rows, each all zero except for a single randomly chosen column. Each measurement is corrupted by AWGN at the 30 dB SNR level and the measurement noise covariance matrix is $\boldsymbol{R}_i = \sigma_v^2 \boldsymbol{I}$ where σ_v^2 is the noise variance. The initial state mean is $\boldsymbol{\mu}_1 = \boldsymbol{0}$ and the initial state covariance $\boldsymbol{\Pi}_1$ is a Toeplitz matrix with three bands indicating correlation with immediate neighbors only. The state transition operator is $\boldsymbol{F}_i = \boldsymbol{I}$, which models the state evolution as a random walk, and the state noise covariance $\boldsymbol{Q}_i$ is a Toeplitz matrix with three bands. Finally, the covariance taper matrices $\boldsymbol{C}_i$, $\boldsymbol{C}'_i$, and $\boldsymbol{C}''_i$ are also a Toeplitz matrix with three bands.

The results of the experiment are depicted in Figure 1 which shows the output of the KS, LKS, and EnKS when the ensemble size is $L = 16$. The qualitative similarity between the KS and LKS results demonstrates that the covariance taper does not introduce a large bias in this example. Finally, the similarities between the ensemble method (EnKS) and the limiting solutions (LKS) demonstrate that the relatively small ensemble size of $L = 16$ is sufficient to obtain qualitatively faithful estimates in this example. Quantitatively, the relative error between the ground truth ($\boldsymbol{x}_i$ for $1 \leq i \leq I$) and the KS ($\widehat{\boldsymbol{x}}_{i|1:I}$), LKS ($\widetilde{\boldsymbol{x}}^{\infty}_{i|1:I}$), and EnKS ($\widetilde{\boldsymbol{x}}_{i|1:I}$) estimates is 0.193, 0.248, and 0.291, respectively.

8. CONCLUSIONS

This paper has developed a new localized ensemble Kalman smoother. The theoretical proof of convergence shows that the approach has a well defined limit as the ensemble size increases and the method is a Monte Carlo approximation to the localized Kalman smoother. A connection to spatial-temporal adaptive processing from the radar literature provides the means to implement a numerically stable ensemble Kalman filter and smoother. The numerical example has shown the differences between the LMMSE optimal Kalman smoother, the biased but deterministic localized Kalman smoother, and the ensemble Kalman smoother, the Monte Carlo approximation to the localized Kalman smoother.

9. REFERENCES

[1] G. Evensen, "The ensemble Kalman filter: Theoretical formulation and practical implementation," *Ocean Dynam.*, vol. 53, pp. 343–367, 2003.

[2] M. D. Butala, R. J. Hewett, R. A. Frazin, and F. Kamalabadi, "Dynamic three-dimensional tomography of the solar corona," *Solar Phys.*, vol. 262, pp. 495–509, 2010.

[3] S. P. Khare, J. L. Anderson, T. J. Hoar, and D. Nychka, "An investigation into the application of an ensemble Kalman smoother to high-dimensional geophysical systems," *Tellus*, vol. 60A, pp. 97–112, 2008.

[4] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Upper Saddle River, NJ: Prentice-Hall, 2000.

[5] M. D. Butala, R. A. Frazin, Y. Chen, and F. Kamalabadi, "Tomographic imaging of dynamic objects with the ensemble Kalman filter," *IEEE Trans. Image Process.*, vol. 18, pp. 1573–1587, 2009.

[6] J. R. Guerci, *Space-Time Adaptive Processing for Radar*. Norwood, MA: Artech House, 2003.

[7] C. R. Rao and M. B. Rao, *Matrix Algebra and Its Applications to Statistics and Econometrics*. Singapore: World Scientific Publishing Company, 1998.